NOTE ON REPEATED SELECTION IN THE NORMAL CASE

Technical Report No. 19

Department of Navy

Office of Naval Research

Contract No. Nonr-409(39)
Project No. (NR 042-212)

BIOMETRICS UNIT
DEPARTMENT OF PLANT BREEDING

NEW YORK STATE COLLEGE OF AGRICULTURE

CORNELL UNIVERSITY

ITHACA, NEW YORK

NOTE ON REPEATED SELECTION IN THE
NORMAL CASE

Technical Report No. 19

Department of Navy

Office of Naval Research

Contract No. Nonr-409(39)
Project No. (NR 042-212)

D. S. Robson

Biometrics Unit

New York State College of Agriculture

Cornell University

Ithaca, New York

This work was supported in part by the Office of Naval Research.

NOTE ON REPEATED SELECTION IN THE NORMAL CASE

BU-166-M D. S. Robson April, 1964

ABSTRACT

A k-cycle selection model is specified by a (k+1)-variate normal distribution of the variables $X, Y_1 = X+\epsilon_1, \cdots, Y_k = X+\epsilon_k$ with selection at the i^{th} stage removing a fraction

$$P_i = P(Y_i > y_i \mid Y_1 > y_1, \cdots, Y_{i-1} > y_{i-1})$$

The distribution of X in this selected fraction is then convolved with the $N(0, \mathcal{E}^2_{i+1})$ distribution of ϵ_{i+1} to form the distribution of Y_{i+1}. An expression is given for the characteristic function of X in the k^{th} selected fraction.

Biometrics Unit, Plant Breeding Department, Cornell University

NOTE ON REPEATED SELECTION IN THE NORMAL CASE

D. S. Robson

BU-166-M April, 1964

Selection for a quantitative trait often continues for several cycles, as in the successive annual screening of a plant population in the process of developing new varieties. With plant selection, as with most other selection problems, the trait x being selected for cannot be measured without error, and actual selections are based on the observation $y_i = x + e_i$ in the i^{th} cycle of the process. We shall assume here that the error chance variable e_i is $N(0,\varepsilon_i^2)$ (normally distributed with mean 0 and variance ε_i^2) and that the error e_i attaching to x in the i^{th} stage is independent of the error e_j attaching to that same x (or any other x) in the j^{th} stage. Further, we suppose that in the unselected population the chance variable x is $N(\xi,\sigma^2)$, so that $y_1=x+e_1$ is $N(\xi,\omega_1^2 = \sigma^2 + \varepsilon_1^2)$.

The population available at the k^{th} stage is assumed to be of infinite size, and selection consists of removing the upper fraction P_k of the available y-population for further selection at stage k+1. The fraction of the original population available for selection at stage k+1 is therefore $P_1P_2\cdots P_k$, and our concern here shall lie with the distribution of x in this remaining fraction. These fractions are defined by

$$P_1 = P(Y_1 > y_1)$$

$$P_1P_2 = P_1P(Y_2 > y_2 \mid Y_1 > y_1)$$

$$P_1P_2\cdots P_k = P_1P_2\cdots P_{k-1}P(Y_k > y_k \mid Y_1 > y_1, Y_2 > y_2, \cdots, Y_{k-1} > y_{k-1})$$

and our results are based upon the observation that this remaining fraction is

simply the tail probability in a k-variate normal distribution,

$$P_1P_2\cdots P_k = P(Y_1 > y_1, Y_2 > y_2, \cdots, Y_k > y_k)$$

Since the joint distribution of $X, Y_1, Y_2, \cdots, Y_k$ is the (k+1)-variate normal distribution with mean ξ and covariance matrix

$$\Lambda = \begin{bmatrix} \sigma^2 & \sigma^2 & \cdots & \sigma^2 \\ \sigma^2 & \omega_1^2 & \cdots & \sigma^2 \\ \vdots & \cdot & & \\ \sigma^2 & \sigma^2 & \cdots & \omega_k^2 \end{bmatrix} = [\sigma_{ij}]$$

then the distribution of x for fixed values of $Y_1, \cdots, Y_k$ is normal with mean

$$E(X|y_1,\cdots,y_k) = \xi - \frac{\Lambda_{o1}}{\Lambda_{oo}}(y_1-\xi) - \cdots - \frac{\Lambda_{ok}}{\Lambda_{oo}}(y_k-\xi)$$

and

$$\mathrm{var}(X|y_1,\cdots,y_k) = \frac{\Lambda}{\Lambda_{oo}}$$

where Λ is the determinant of Λ and Λ_{ij} is the cofactor of the ij^{th} element of Λ.

The joint distribution of $Y_1, \cdots, Y_k$ is normal with mean $\xi = (\xi, \cdots, \xi)$ and covariance matrix Λ_{oo}. Using the expansion

$$\Lambda = \sigma_{oo}\Lambda_{oo} - \sum_{i,j=1}^{k} \sigma_{io}\sigma_{oj}\Lambda_{oo\cdot ij}$$

where $\Lambda_{oo\cdot ij}$ is the cofactor of σ_{ij} in Λ_{oo}, we may then express the conditional

moment generating function of X as

$$E(e^{tX}|y_1,\cdots,y_k)$$

$$= e^{t\xi + \frac{t^2}{2}(\sigma_{oo} - \frac{1}{\Lambda_{oo}} \sum_{i,j=1}^{k} \sigma_{io}\sigma_{oj}\Lambda_{oo\cdot ij}) + \frac{t}{\Lambda_{oo}} \sum_{i,j=1}^{k} \sigma_{io}(y_j-\xi)\Lambda_{oo\cdot ij}}$$

and then

$$E(e^{tX}|Y_1 > y_1,\cdots,Y_k > y_k) = e^{t\xi + \frac{t^2}{2}\sigma_{oo}} (P_1P_2\cdots P_k)^{-1} \frac{1}{(2\Pi)^{\frac{k}{2}}\sqrt{\Lambda_{oo}}} \cdot$$

$$\int_{\{u_i > y_i\}} e^{-\frac{1}{2\Lambda_{oo}} \sum_{i,j=1}^{k} [(u_i-\xi)(u_j-\xi)-2t\sigma_{io}(u_j-\xi)+t^2\sigma_{io}\sigma_{oj}]\Lambda_{oo\cdot ij}} du_1\cdots du_k$$

The exponent in the integral reduces to

$$\sum_{i,j=1}^{k} (u_i-\xi-\sigma_{io}t)(u_j-\xi-\sigma_{oj}t)\Lambda_{oo\cdot ij}$$

hence, transforming to the standard normal $z_i = (y_i-\xi)/\sqrt{\sigma_{ii}}$, we obtain

$$E(e^{tX}| \frac{Y_1-\xi}{\omega_1} > z_1,\cdots, \frac{Y_k-\xi}{\omega_k} > z_k)$$

$$= \frac{e^{t\xi + \frac{t^2}{2}\sigma^2}}{P_1P_2\cdots P_k} \frac{1}{(2\Pi)^{\frac{k}{2}}\sqrt{R_{oo}}} \int_{\{v_i > z_i - \frac{\sigma^2}{\omega_i}t\}} e^{-\frac{1}{2P_{oo}} \sum_{i,j=1}^{k} R_{oo\cdot ij}v_iv_j} dv_1\cdots dv_k$$

where

$$R_{oo} = \begin{vmatrix} 1 & \frac{\sigma^2}{\omega_1\omega_2} & \cdots & \frac{\sigma^2}{\omega_1\omega_k} \\ \frac{\sigma^2}{\omega_1\omega_2} & 1 & \cdots & \frac{\sigma^2}{\omega_2\omega_k} \\ \vdots & \vdots & & \vdots \\ \frac{\sigma^2}{\omega_1\omega_k} & \frac{\sigma^2}{\omega_2\omega_k} & \cdots & 1 \end{vmatrix}$$

The mean value of X in this selected fraction of the population is obtained by differentiating once with respect to t, first writing

$$E(e^{tX} \mid \frac{Y_1-\xi}{\omega_1} > z_1, \cdots, \frac{Y_k-\xi}{\omega_k} > z_k)$$

$$= \varphi_X(t) P_{R_{oo}}(v_1 > z_1 - \frac{\sigma^2}{\omega_1} t, \cdots, v_k > z_k - \frac{\sigma^2}{\omega_k} t)/P_1 P_2 \cdots P_k$$

so that the derivative becomes

$$\frac{1}{P_1 P_2 \cdots P_k} \{\varphi_X'(t) P_{R_{oo}}(v_1 > z_1 - \frac{\sigma^2}{\omega_1} t, \cdots, v_k > z_k - \frac{\sigma^2}{\omega_k} t)$$

$$+ \frac{\sigma^2}{\sqrt{2\Pi}} \sum_{j=1}^{k} \frac{1}{\omega_j} e^{-\frac{1}{2}(z_j - \frac{\sigma^2}{\omega_j} t)^2}$$

$$P_{R_{oo}}(v_1 > z_1 - \frac{\sigma^2}{\omega_1} t, \cdots, v_{j-1} > z_{j-1} - \frac{\sigma^2}{\omega_{j-1}} t, v_{j+1} > z_{j+1}$$

$$- \frac{\sigma^2}{\omega_{j+1}} t, \cdots, v_k > z_k - \frac{\sigma^2}{\omega_k} t | z_j)\}$$

Setting t=0, we obtain the mean value

$$\xi + \frac{\sigma^2}{P_1 P_2 \cdots P_k} \sum_{j=1}^{k} \frac{1}{\omega_j \sqrt{2\Pi}} e^{-\frac{z_j^2}{2}} P_{R_{oo}}(v_1 > z_1, \cdots, v_{j-1} > z_{j-1}, v_{j+1} > z_{j+1}, \cdots,$$

$$v_k > z_k | v_j = z_j)$$

or

$$\xi + \frac{\sigma^2}{P_1 P_2 \cdots P_k} \sum_{j=1}^{k} \frac{1}{\omega_j \sqrt{2\Pi}} e^{-\frac{z_j^2}{2}} P_{R_{oo}^{(j)}} (u_1 > z_1 - \frac{\sigma^2}{\omega_1 \omega_j} z_j, \cdots, u_{j-1} > z_{j-1}$$

$$- \frac{\sigma^2}{\omega_{j-1}\omega_j} z_j, u_{j+1} > z_{j+1} - \frac{\sigma^2}{\omega_j \omega_{j+1}} z_j, \cdots, u_k > z_k - \frac{\sigma^2}{\omega_j \omega_k} z_j)$$

where

$$R_{00}^{(j)} = \begin{vmatrix} 1-\frac{\sigma^4}{\omega_1^2\omega_j^2} & \cdots & \frac{\sigma^2}{\omega_1\omega_{j-1}}\left(1-\frac{\sigma^2}{\omega_j^2}\right) & \frac{\sigma^2}{\omega_1\omega_{j+1}}\left(1-\frac{\sigma^2}{\omega_j^2}\right) & \cdots & \frac{\sigma^2}{\omega_1\omega_k}\left(1-\frac{\sigma^2}{\omega_j^2}\right) \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\sigma^2}{\omega_1\omega_{j-1}}\left(1-\frac{\sigma^2}{\omega_j^2}\right) & \cdots & 1-\frac{\sigma^4}{\omega_{j-1}^2\omega_j^2} & \frac{\sigma^2}{\omega_{j-1}\omega_{j+1}}\left(1-\frac{\sigma^2}{\omega_j^2}\right) & \cdots & \frac{\sigma^2}{\omega_{j-1}\omega_k}\left(1-\frac{\sigma^2}{\omega_j^2}\right) \\ \frac{\sigma^2}{\omega_1\omega_{j+1}}\left(1-\frac{\sigma^2}{\omega_j^2}\right) & \cdots & \frac{\sigma^2}{\omega_{j-1}\omega_{j+1}}\left(1-\frac{\sigma^2}{\omega_j^2}\right) & 1-\frac{\sigma^4}{\omega_j^2\omega_{j+1}^2} & \cdots & \frac{\sigma^2}{\omega_{j+1}\omega_k}\left(1-\frac{\sigma^2}{\omega_j^2}\right) \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\sigma^2}{\omega_1\omega_k}\left(1-\frac{\sigma^2}{\omega_j^2}\right) & \cdots & \frac{\sigma^2}{\omega_{j-1}\omega_k}\left(1-\frac{\sigma^2}{\omega_j^2}\right) & \frac{\sigma^2}{\omega_{j+1}\omega_k}\left(1-\frac{\sigma^2}{\omega_j^2}\right) & \cdots & 1-\frac{\sigma^4}{\omega_j^2\omega_k^2} \end{vmatrix}$$

CORNELL UNIVERSITY - BIOMETRICS UNIT

Distribution List for Unclassified Technical Reports

Contract Nonr-401(39)
Project (NR 042-212)

Head, Logistics and Mathematical Statistics Branch Office of Naval Research Washington, D. C. 20360	3 copies
Commanding Officer Office of Naval Research Branch Office Navy 100 Fleet Post Office New York, New York	2 copies
Defense Documentation Center Cameron Station Alexandria, Virginia 22314	20 copies
Defense Logistics Studies Information Exchange Army Logistics Management Center Fort Lee, Virginia Attn: William B. Whichard	1 copy
Technical Information Officer Naval Research Laboratory Washington, D. C. 20390	6 copies
Commanding Officer Office of Naval Research Branch Office 207 West 24th Street New York, New York 10011 Attn: J. Laderman	1 copy
Commanding Officer Office of Naval Research Branch Office 1030 East Green Street Pasadena 1, California Attn: Dr. A. R. Laufer	1 copy
Bureau of Supplies and Accounts Code OW Department of the Navy Washington 25, D. C.	1 copy
Institute for Defense Analyses Communications Research Division von Newmann Hall Princeton, New Jersey	1 copy

University of Chicago
Statistical Research Center
Chicago, Illinois
Attn: Prof. Paul Meier

Stanford University
Applied Mathematics & Statistics Lab.
Stanford, California
Attn: Prof. Gerald J. Lieberman

Florida State University
Department of Statistics
Tallahassee, Florida
Attn: Dr. Ralph A. Bradley

Princeton University
Department of Mathematics
Princeton, New Jersey

Columbia University
Department of Mathematical Statistics
New York 27, New York
Attn: Prof. T. W. Anderson

University of California
Department of Statistics
Berkeley 4, California
Attn: Prof. J. Neyman

University of Washington
Department of Mathematics
Seattle 5, Washington
Attn: Prof. Z. W. Birnbaum

Cornell University
Department of Mathematics
Ithaca, New York
Attn: Prof. J. Wolfowitz

Harvard University
Department of Statistics
Cambridge, Massachusetts
Attn: Prof. W. G. Cochran

Florida State University
Department of Statistics
Tallahassee, Florida
Attn: Prof. I. R. Savage

Columbia University
Department of Industrial Engineering
New York 27, New York
Attn: Prof. Cyrus Derman

Columbia University
Department of Mathematics
New York 27, New York
Attn: Prof. H. Robbins

New York University
Institute of Mathematical Sciences
New York 3, New York
Attn: Prof. W. M. Hirsch

Cornell University
Department of Plant Breeding
Biometrics Unit
Ithaca, New York
Attn: Walter T. Federer

University of North Carolina
Statistics Department
Chapel Hill, North Carolina
Attn: Prof. Walter L. Smith

Michigan State University
Department of Statistics
East Lansing, Michigan
Attn: Prof. Herman Rubin

Brown University
Division of Applied Mathematics
Providence 12, Rhode Island
Attn: Prof. M. Rosenblatt

New York University
Department of Industrial Engineering
and Operations Research
Bronx 63, New York
Attn: Prof. J. H. Kao

University of Wisconsin
Department of Statistics
Madison, Wisconsin
Attn: Prof. G. E. P. Box

The Research Triangle Institute
Statistics Research Division
505 West Chapel Hill Street
Durham, North Carolina
Attn: Dr. Malcolm R. Leadbetter

Institute of Mathematical Statistics
University of Copenhagen
Copenhagen, Denmark
Attn: Prof. Anders Hald

University of Michigan
Department of Mathematics
Ann Arbor, Michigan
Attn: Prof. L. J. Savage

Massachusetts Institute of Technology
Department of Electrical Engineering
Cambridge, Massachusetts
Attn: Dr. R. A. Howard

The Johns Hopkins University
Department of Mathematical Statistics
34th & Charles Streets
Baltimore 18, Maryland
Attn: Prof. Geoffrey S. Watson

Stanford University
Department of Statistics
Stanford, California
Attn: Prof. E. Parzen

Arcon Corporation
803 Massachusetts Avenue
Lexington 73, Massachusetts
Attn: Dr. Arthur Albert

University of California
Institute of Engineering Research
Berkeley 4, California
Attn: Prof. R. E. Barlow

Michigan State University
Department of Statistics
East Lansing, Michigan
Attn: Prof. J. Gani

Rocketdyne - A Division of North
American Aviation, Inc.
6633 Canoga Avenue
Canoga Park, California
Attn: Dr. Roy Goodman
Dr. Robert Zimmerman

Rutgers - The State University
Statistics Center
New Burnswick, New Jersey

Yale University
Department of Statistics
New Haven, Connecticut
Attn: Prof. F. J. Anscombe

Purdue University
Division of Mathematical Sciences
Lafayette, Indiana
Attn: Prof. S. S. Gupta

Cornell University
Department of Industrial Engineering
Ithaca, New York
Attn: Prof. Robert Bechhofer

Stanford University
Department of Statistics
Stanford, California
Attn: Professor C. Stein

Applied Mathematics and Statistics Lab.
Department of Statistics
Stanford University
Stanford, California
Attn: Prof. H. Solomon